

The Manchester School Vol 00 No. 00 00–00 Month 2016
doi: 10.1111/manc.12147

PERFORMANCE-RELATED PAY, EFFICIENCY WAGES AND THE SHAPE OF THE TENURE-EARNINGS PROFILE*

by

JOHN G. SESSIONS

University of Bath

and

JOHN D. SKÅTUN†

University of Aberdeen

Whilst it is well known that performance-related pay (PRP) may increase wage inequality within a firm, there is an inter-temporal life-cycle aspect that has been largely ignored in the literature. In this paper, we investigate theoretically how the introduction of PRP will influence the wage and remuneration profile over time. We develop a simple two-period model of efficiency wages that rationalizes recent empirical findings suggesting PRP flattens the pay-tenure profile. Such attenuation has important implications for the credibility of long-term employment contracts as it suggests that agency rather than human capital considerations drive the profile.

1 INTRODUCTION

Efficiency-wage theory predicts that firms can raise worker productivity by adopting a carrot and stick approach of paying a supra-competitive wage, devoting resources to monitoring, and dismissing any workers it detects as shirking (Shapiro and Stiglitz, 1984).¹ By extension, if employment contracts are multi-period then firms can raise worker productivity by *tilting* the remuneration package over time, paying workers less than the value of their marginal product when they are relatively short-tenured, and correspondingly more than the value of their marginal product when they are relatively long-tenured. Such a pay profile provides workers with *ex post* rents that they will be reluctant to jeopardize. If reducing effort increases the probability of involuntary termination, then upward sloping pay profiles raise the cost of shirking and induce workers to raise effort.²

* Manuscript received 25.1.14; final version received 4.1.16.

† We are grateful to three anonymous referees and the editors of this journal for helpful comments. The normal disclaimer applies.

¹The Shapiro-Stiglitz model regards worker effort as synonymous with productivity. This need not be the case. Other conduits through which efficiency wages might impact upon productivity include reduced turnover (Salop, 1979; Stiglitz, 1985), adverse selection (Weiss, 1980) and worker morale (Akerlof, 1982).

²Recent support for this prediction is offered by Adams and Heywood (2011) who find evidence from both US and Australian data that deferred compensation, whether in the form of

An alternative method of raising worker productivity is to divest a share of the firm into the hands of workers through collective (e.g. profit sharing, employee share ownership) and/or individual performance-related pay (PRP) schemes (see Blinder, 1990; Weitzman and Kruse, 1990). Such schemes, especially individual ones, directly reduce the marginal benefit of shirking and, therefore, have implications for both the stationary and dynamic versions of the efficiency wage story. In terms of the latter, introspection would suggest—and empirical work has confirmed—that the greater the component of total pay that is derived from PRP, then the flatter is the slope of the pay-tenure profile *ceteris paribus* (see, for example, Lazear and Moore, 1984; Brown and Sessions, 2006).

In this paper, we compare and contrast a pure efficiency wage setup with a mixed PRP framework and rationalize formally the relationship between remuneration structures and the nature of the pay-tenure profile. Under reasonable assumptions, we find that effort and remuneration exhibit a greater tenure gradient under a wage-only contract as compared to contracts where remuneration includes some element of PRP. With pay in wage-only contracts being higher later on in the firm-worker relationship, they are also compensatingly lower earlier on as compared with firm-worker contracts that include PRP. The results of the simple model presented below are therefore compatible with, and lend theoretical support to, previous empirical findings.

Whilst we emphasize the efficiency wage nature of our model, the study of involuntary unemployment outcomes is not our main focus. As such, we deviate from the Shapiro and Stiglitz tradition by not endogenizing unemployment rates.³ There are two reasons for this: First, it allows us to concentrate on our main objective of studying payment profiles over time and across different payment structures; and second, the bonding critique of efficiency wages suggests that involuntary unemployment does not arise with up-front payments (see Carmichael, 1990). Though we rule out explicit bond payments, a result of the model has remunerations rising over time, suggesting that younger workers pay an entrance fee to gain higher future payments. Thus, the involuntary unemployment consequences remain far from clear.

Finally, our model relates to the separate literature that PRP may influence and increase wage inequality within a firm (see, for example, Lemieux *et al.*, 2009). Whilst we do not dispute that this may be the case in any particular period, there is also an inter-temporal lifecycle aspect that has been largely ignored. Our purpose is to investigate how the introduction of PRP influences the wage and remuneration profile over time. This approach is as

steeper tenure-wage profiles or pensions, is associated with higher self-reported worker effort, with the increase in effort declining as the chance of job separation rises.

³A recent paper in this tradition is Basu and Felkey (2009) who find multiple unemployment equilibria.

such novel. We do not seek to determine why and when PRP is paid - see in this respect, for example, Lazear (2000b), who discusses the use of PRP as a sorting mechanism when the workforce is heterogeneous, and Gibbons (2005), who highlights the perennial conflict between offering a fixed wage for insurance purposes and a variable PRP element to generate the correct incentives. Our paper is in some aspects similar to MacLeod and Malcomson (1998), who in the one sector version of their model find that a fixed (efficiency) wage dominates when there is unemployment, but that PRP dominates when qualified workers are in short supply. We, in contrast, do not seek to determine whether or not PRP dominates a within-period fixed wage but rather, by appealing to realism, presume that fixed wages and PRP coexist and thereby investigate how the presence of exogenously determined PRP might compress life cycle remuneration.

The paper is set out as follows: Section 2 discusses the wage-seniority nexus whilst Section 3 sets out our modelling framework. Section 4 investigates how the supply of effort depends on pay under 'wage-only' and 'PRP' schemes, the latter being defined as schemes that comprise both a wage-only and performance dependent term.⁴ Section 5 investigates the demand side decisions by the firm and Section 6 offers some concluding comments.

2 THE WAGE-SENIORITY NEXUS

The positive correlation between seniority and pay is one of the most robust empirical findings in labour economics—for surveys of the theoretical and the empirical literature see Hutchens (1989), Carmichael (1990), Polachek and Siebert (1992), and Lazear (2000a).⁵ The source of the relationship, however, is somewhat ambiguous.

The conventional explanation until the 1980s for the relationship was that earnings reflected the acquisition of, and reward to, general and specific human capital. Workers became more productive, and hence better remunerated, over time because of investments in training. Workers paid for general training, and subsidized specific training, by accepting early career (i.e. training) wages below the value of their marginal product to the firm. Latter career (i.e. trained) wages reflected the increase in worker productivity; fully, in the case of general training, and partially, in the case of specific training. In either case, an upward sloping wage profile emerges; wages increase with

⁴We use the term 'contract' in a loose sense and interchangeable with 'payment arrangement', where the firm does not pre-commit to a particular wage/remuneration profile. Any ensuing payment profiles are rather a reflection of remuneration policies which are set on a single-period basis but where, in a two-period setting, period-one decisions are contingent on the realized pay of period-two.

⁵The finding is not completely unchallenged—see, for example, Hilmer (2011), Altonji and Williams (2005) and Dostie (2005).

seniority because productivity increases with seniority (Mincer, 1958; Becker, 1962; Ben-Porath, 1967).

The human capital explanation was challenged in a series of papers by Lazear (1979, 1981, 1983) and Medoff and Abraham (1980, 1981). Lazear observes that mandatory retirement and actuarially unfair pension schemes that encouraged early retirement were incompatible with human capital theory. Why would firms establish human resource policies whereby an employee is paid, and thus evidently valued, today but then either forced or induced to quit tomorrow? Such policies contradict the human capital thesis that senior workers are paid no more than their marginal product, particularly when wages can be adjusted downwards if productivity declines with age.⁶

Lazear reconciles the various phenomena by focussing on contracts that discourage employee shirking and other malfeasance over an employee's life cycle in situations where monitoring worker effort is problematic. The basic idea is that workers and firms enter into contracts in which pay is less than the value of the workers' marginal product early on in their job tenure, and correspondingly more than the value of their marginal product in their later years. By back-loading compensation workers receive *ex post* rents that they are reluctant to lose. The upward sloping wage profiles increase the cost of shirking and encourage workers to raise effort.

Lazear's explanation cuts the link between productivity and pay; wages grow with seniority irrespective of productivity. And whilst it makes sense for the firm to pay wages in excess of the value of a worker's marginal product for a period of time, it is not sensible to do this indefinitely. There will come a point when the present discounted value of the worker's marginal product equals the present discounted value of his remuneration package. This would imply, from the firm's perspective, an optimal retirement date and a need for policies to force or encourage the worker's retirement.

Several studies have attempted to discriminate empirically between the agency and the human capital explanations. Hutchens (1987), for example, focuses on the implicit trade-off between the use of deferred payment contracts and the difficulty of monitoring and finds that Lazear-type characteristics (i.e. wage profiles, mandatory retirement, pension schemes, long job tenures) tend not to be associated with jobs that are conducive to monitoring. Similarly, Sessions and Theodoropoulos (2013), using matched British employer–employee data, find a negative relationship between the slope of the wage-tenure profile and the level of monitoring.

⁶Medoff and Abraham (1980, 1981) highlighted a related conundrum in their analysis of data on pay and supervisor performance ratings. They found that although relative performance ratings within a particular job grade did not increase with experience in the job grade, relative pay did. Again, such a finding is incompatible with the human capital position that earnings increase with seniority because productivity increases with seniority.

Lazear and Moore (1984), in a US study, address the issue by considering the empirical evidence regarding the relative ‘flatness’ of self-employed workers’ wage profiles (Wolpin, 1977; Fuchs, 1981). Such a finding is puzzling as investments in physical capital would tend to depress observed wages for the early career self-employed, whilst subsequent returns to those investments would tend to raise observed future wages. Both factors imply that the wage profiles of self-employed workers should be steeper than those of otherwise similar wage and salary workers. Lazear and Moore (1984) rationalize the finding by highlighting the duality of principal and ownership intrinsic to self-employment. Observed wage profiles, they argue, reflect a disharmony of interests prevalent in the employment relation, a dissonance that is, by definition, absent from self-employment. By steepening the wage profile, employers are able to induce their employees to work harder. The self-employed require no such internal incentive mechanism and thus may be used as a control group to test the theoretical prior that the profile is determined primarily by agency as opposed to human capital considerations.

The UK study by Brown and Sessions (2006) then generalizes Lazear and Moore’s approach by comparing the earnings profiles of self-employed workers, wage/salary workers, and workers employed under PRP schemes. If agency considerations are important in driving the earnings profile, and if PRP workers face an intermediate degree of agency as compared to wage/salary and self-employed workers, then the earnings of PRP workers should increase at an intermediate rate with tenure as compared to these other types of workers.

Both Lazear and Moore (1984) and Brown and Sessions (2006) find convincing empirical evidence that it is agency considerations that drive the pay-tenure profile. Related studies include Flabbi and Ichino (2001), Dohmen (2004) and Zwick (2011), all of whom find evidence from firm personnel records to replicate Medoff and Abraham’s (1980, 1981) finding that earning increases within firms do not reflect increases in productivity. Experimental evidence supporting an agency driven profile is found by Huck *et al.* (2011). In what follows, we endeavour to formally rationalize these various findings.

3 THE MODELLING FRAMEWORK

We assume a dual-labour market setting comprising a primary-sector offering high paid jobs in which workers have some discretion over effort, and who therefore require monitoring, and a secondary-sector offering low paid menial jobs in which shirking is impossible (Bulow and Summers, 1986). Our focus is the primary sector and our aim in the following two-period model is to illustrate the relationship there between agency, worker ‘equity’ (i.e. the extent of any PRP) and the nature of the earnings profile. For ease

of analytical exposition we abstract from considerations regarding human capital and focus instead on the supply and demand aspects of cost-minimising contracts offered by a firm in the presence of asymmetric information regarding worker effort.⁷ Two regimes are considered: a 'wage-only' regime, where firms can only use fixed wages to compensate workers; and a PRP regime where remuneration comprises two elements; a fixed wage and an additional component related to individual worker performance. We assume that both regimes offer single period (i.e. 'spot') or two-period (i.e. 'lifetime') employment contracts. This section outlines the modelling aspects common to both regimes.

3.1 The Informational Context

To help conceptualize the informational context, consider the time sequence for each period t illustrated in Fig. 1 following:

The sequence is common to both regimes outlined above. In Stage 1, for each period t , an effort level is chosen by the worker but is not necessarily known by the firm. After effort is chosen the worker experiences a shock to his productivity in Stage 2, which the firm cannot observe. Output, however, is common knowledge in Stage 3. The firm has then a probability p of observing the effort of the worker in Stage 4. Finally, in Stage 5, a non-shirker or undetected shirker is remunerated, whilst a detected shirker is laid-off.

The key difference between the wage-only and PRP regimes is that PRP remuneration in Stage 5 is contingent upon the output realized in Stage 3. This output is itself a function of worker effort and the idiosyncratic shock to worker productivity in the first two stages. Thus, exerting effort benefits the worker in two ways under PRP. It makes remuneration more likely in Stage 5 and also, as PRP is dependent on production *ceteris paribus*, increases the remuneration payment through higher average output levels in Stage 3. This is in contrast to the wage-only case where the output signal of Stage 3 is not used in determining payment. Nevertheless, a higher non-shirking effort in the wage-only case still guarantees remuneration in Stage 5, as is the case under PRP. Similar in both regimes is the role remuneration plays to make the worker exert effort. Indeed, under both regimes the firm seeks to find the optimal incentive compatible payment schedule that maximizes profit. This is further outlined in the following sections of the paper where we show that, though higher remuneration implies higher effort levels under both regimes, there are several important distinguishing differences between the wage-only and the PRP regime.

⁷Thus, any wage growth must be generated solely from agency reasons. Were one to introduce human capital into the model then wage growth could emanate from two sources; agency considerations and human capital accumulation.

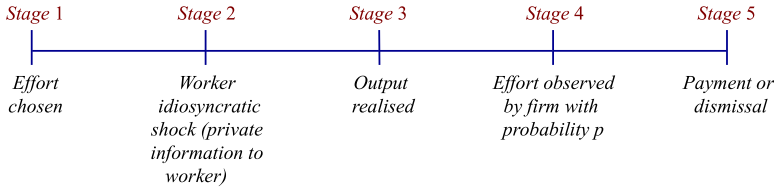


FIG. 1. The Informational Context in Each Time Period

3.2 Output and Effort

We retain the common assumption in the literature that the employer receives a noisy signal of worker output but is unable to ascertain effort directly. The key critical difference between the two payment structures is that PRP remuneration is contingent on the signal whereas wage-only remuneration is not.

To be sure, each worker has a stochastic revenue function, $y_{it} = \theta_{it}f(e_t)$, denoting the revenue associated with his job in state i at time t as a function of effort. We assume that $df(e_t)/de_t \equiv f'(e_t) > 0$ and $d^2f(e_t)/de_t^2 \equiv f''(e_t) < 0$ such that output increases at a diminishing rate with effort. The shift-parameter, θ_{it} , represents a random shock to productivity in state i at time t and is distributed between two values: a lower value θ_L , bounded at zero, and an upper value $\theta_H > \theta_L = 0$. For an individual worker, θ_i reflects relative misfortune (when it is low) or luck (when it is high). For its part the firm is able to observe worker revenue, but is unable to observe either worker effort, e_t , or 'luck', θ_{it} .

3.3 Worker and Firm Objectives

Workers are identical, risk neutral and endowed with a working life of two periods. Their separable, periodic utility function is given by $u_t = m_t - g(e_t)$, $t = 1, 2$, where m_t denotes the worker's remuneration in period t . Remuneration in the wage-only regime is simply a fixed wage \bar{w}_t . Under PRP, total remuneration, \hat{w}_t^T , comprises both a fixed wage component, \bar{w}_2 , and a variable, performance related output payment with the parameter $\lambda \in (0, 1)$ reflecting the relative weight attributed to the latter vis. $\hat{w}_t^T = (1 - \lambda)\bar{w}_t + \lambda\theta_{it}f(\hat{e}_t)$.⁸ It is apparent that a wage-only contract is a special of PRP in which $\lambda = 0$ such that $\hat{w}_t^T = \bar{w}_t = \bar{w}_t$.

⁸We assume in what follows that the extent of PRP, as measured by λ , is exogenous. This is obviously a simplistic assumption and a more complete exposition would seek to explain the distribution of different contractual arrangements. Our main focus is nevertheless not on the causes of different payment schemes but rather on their effect on remuneration profiles over time. For a theoretical investigation that focuses on the former see MacLeod and Malcomson (1998) who show that PRP dominates efficiency wages when vacancy costs are low and there are few available qualified workers. For empirical evidence see Ortín-Angel and Salas-

In both regimes $g(\cdot)$ is a continuous cost of effort function with $g(0)=0$ such that no cost is incurred at zero effort. To ensure that the worker's maximisation problem is well behaved we make the usual assumption that cost is increasing and convex in effort vis. $dg(e_t)/de_t \equiv g'(e_t) > 0$ and $d^2g(e_t)/de_t^2 \equiv g''(e_t) > 0$. We also assume that the (identical) workers either supply the firm's required level of effort or shirk by providing less than the level required.⁹

The firm's objective is to maximize per-worker profit, $\pi_{it}=y_{it}-m_t$. Whilst able to observe worker revenue, the firm is unable to observe either effort or luck and is therefore never able to positively identify a shirker.¹⁰ The firm is thus obliged to monitor workers as a deterrent and we reflect this in a probability $p \in (0, 1)$ of shirking behaviour being detected.¹¹ We assume that detection of shirking implies instantaneous dismissal. Under a spot contract, or in the second period of a lifetime contract, this entitles the worker to claim unemployment utility $b_t=b > 0$ for that period. If caught shirking in the first period of a lifetime contract then the worker is able to claim unemployment benefit for period-one and would then face an expected period-two utility of $\bar{v}_2=s(u_2)[w_2^a-g(e_2^a)]+[1-s(u_2)]b$, where $s(u_2)$ is the probability of obtaining employment with another firm at an alternate (i.e. 'outside') wage w_2^a , in return for e_2^a effort, as a function of the period-two unemployment rate, u_2 . We assume that the 'employment' function, $s(u_2)$, is decreasing in u_2 such that $s'(u_2) < 0$ with $s(0)=1$ and $s(1)=0$.

We adopt the two-sector framework of Demiralp (2011) in which workers who are dismissed from primary sector firms are unable to regain work in that sector. Instead they face either unemployment or work in the secondary sector where firms perfectly observe effort so that there, by definition, is an absence of the moral hazard problem that arises in the primary sector. With effort observable, the secondary sector firm offers wage contracts just sufficient to induce participation, implying that such workers are indifferent between working and unemployment. As in Demiralp (2011), detected shirkers are held down to their reservation utility by firms such that, in our case, secondary sector period-two pay is given by $w_2^a=b+g(e_2^a)$ which implies

Fumás (1998) who, in the context of our model, find evidence that λ varies across both different sectors and levels of a firm's hierarchy. They also find that bonus payments increase with age, although it is unclear whether this is due to a productivity effect or a reflection of a change in λ across age groups.

⁹Under a wage-only regime there is no benefit to a shirker from supplying any $e > 0$ such that shirking is complete with shirking effort $e=0$. This is not necessarily the case under PRP as there are both costs and benefits from supplying shirking effort $e > 0$ —see Section 4.2 following.

¹⁰A possibility arises that with a sufficiently high realisation of luck the firm may be able to identify 'lucky' non-shirkers. However, to keep matter simple we assume that the firm pays all workers a uniform wage provided that they are not detected shirking.

¹¹The risk of detection, p , reflects the firm's ability and incentive to monitor workers. Whilst we do not model this explicitly, we assume that production and monitoring technologies are such that it is always optimal for the firm to monitor imperfectly.

that $\bar{v}_2 = b$. Intuitively, to ensure primary sector incentive compatibility, secondary sector remuneration has to be lower than the primary sector, equilibrium non-shirking period-two efficiency remuneration under either a wage-only or PRP contract.¹²

To determine optimal remuneration and its relationship with effort, in both the wage-only and the PRP case, we investigate the supply and demand side responses of effort to changes in remuneration. Starting with supply issues, we first determine in Sections 4.1 and 4.2 respectively the incentive compatible remuneration-effort loci under fixed wages (i.e. where $\lambda = 0$) and PRP (i.e. where $0 < \lambda < 1$). After comparing these responses in Section 4.3, and assuming that firms correctly anticipate these supply-side responses, we then use the loci to determine in Sections 5.1 and 5.2 the demand side behaviour reflected in the remuneration that fixed-wage and PRP firms will optimally set. The overall efficiency remuneration/wage therefore incorporates both supply and demand side issues.

4 THE SUPPLY SIDE: INCENTIVE COMPATIBLE EFFORT AND PAY

This section determines the supply relationship between worker effort and remuneration under both the wage-only regime where $\lambda = 0$ and the PRP regime where $\lambda \in (0, 1)$. To distinguish between the two payment regimes, we denote the time dependent level of effort in the wage-only and PRP regimes as \tilde{e}_t and \hat{e}_t respectively. We now turn to investigate the supply side of the two payment schemes in turn.

4.1 Incentive Compatible Remuneration Under a Wage-Only Regime

We investigate first the supply locus of remuneration and effort under wage-only setting. Consider the one-period (spot wage) case. For a given level of effort, \tilde{e}_s , there exists a wage, \tilde{w}_s , at or above which workers will supply (at least) \tilde{e}_s and below which workers will shirk. Paying \tilde{w}_s will, in other words, induce the worker to supply effort \tilde{e}_s . Note that to induce the worker to supply more effort than \tilde{e}_s , the wage would have to rise beyond \tilde{w}_s . Such supply considerations can be summarized by the incentive compatible ‘non-shirking constraint’ (NSC), which specifies the lowest wage required for the worker to supply a given effort level or, equivalently, the maximum effort supplied

¹²There is compelling empirical evidence of unemployment scarring—see for example Arulampalam *et al.* (2001), Gregg (2001), Eliason and Storrie (2006), Biewen and Steffes (2010) and Huttunen *et al.* (2011). Whether such scarring would be sufficient to limit detected shirkers to their reservation utility for the rest of their working life is, however, moot. We make the assumption, which we claim is reasonable in the context of our model, to simplify the exposition. Our focus is on the relative difference in the remuneration profiles between the two regimes and, assuming wage-only and PRP workers have the same outside options, then any such scarring should affect both profiles identically. Note further that unless there is full employment then a detected shirker’s expected period-two utility would be lower than his period-one utility even without scarring.

for a given wage. In the spot market one-period case, the *NSC* under a wage-only contract, denoted NSC_s^w , requires that the effort exerting utility, $\tilde{w}_s - g(\tilde{e}_s)$ must be at least as high as the expected utility of shirking. As shirking is complete under a wage-only contract, this latter is given by the weighted average of unemployment payoff b if detected with probability $p > 0$ and collecting the wage \tilde{w}_s with probability $(1-p) < 1$ vis:

$$NSC_s^w : \tilde{w}_s - g(\tilde{e}_s) \geq pb + (1-p)\tilde{w}_s \quad (1)$$

Satisfaction of NSC_s^w implies the incentive compatible (i.e. efficiency) wage schedule:

$$\tilde{w}_s = b + \frac{g(\tilde{e}_s)}{p} \quad (2)$$

This gives us the required wage at any given level of effort such that workers are just indifferent between shirking and not shirking. This supply schedule holds for a variety of wage and effort levels, the precise combinations of which will be pinned down later when we consider both supply and demand considerations. Note here that the incentive compatible wage level is a function of three elements: It is increasing in terms of the worker's outside unemployment opportunity, b , as the firm will have to pay more to induce effort the better are the worker's alternative employment prospects. Second, the wage is high when the cost of effort, $g(\tilde{e}_s)$, is high as the firm will need to pay the worker more to induce effort the more costly it is for the worker to do so. And third, the wage is lower the higher the probability of detecting a potential shirker, as it is the fear of detection that drives the worker to exert higher effort. Higher detection probabilities thus shade the necessary effort-inducing wage that the firm is compelled to offer.

Now consider the specification of a two-period 'lifetime' wage contract $(\tilde{w}_1, \tilde{w}_2)$. At the start of the second period the firm and the worker are locked together in an employment relationship they know will only last for one more period and as such they face the same effort elicitation considerations as firms and workers contracting in a single-period spot market. Thus, the wage-effort schedule in period-two is naturally and commonly known to be given by expression (2) such that $\tilde{w}_2 = b + [g(\tilde{e}_2)/p]$. Both the firm and worker in period-one are able to correctly anticipate the second period efficiency wage such that the period-one effort compatible inducing wage is dependent on future wages. For example, given any period-one wage the worker will have more to lose if detected shirking and fired in period-one the higher the period-two wage. This in turn lowers the necessary period-one wage required to elicit a particular level of effort.

Formally, the lifetime period-one *NSC* in the wage-only contract with no discounting, denoted NSC_1^w , is given by:

$$NSC_1^w : \tilde{w}_1 + \tilde{w}_2 - g(\tilde{e}_1) - g(\tilde{e}_2) \geq p\beta + (1-p)[\tilde{w}_1 + \tilde{w}_2 - g(\tilde{e}_2)] \quad (3)$$

where $\beta = b + \bar{v}_2$. Undetected shirkers enjoy utility of \tilde{w}_1 now and $\tilde{w}_2 - g(\tilde{e}_2)$ tomorrow, where \tilde{w}_2 is the incentive compatible wage schedule that ensures workers do not shirk in period-two but instead exert the required period-two effort. Detected shirkers receive b in period-one and can expect utility of $\bar{v}_2 < \tilde{w}_2$ in period-two. Satisfaction of NSC_1^W together with the period-two version of (2), $\tilde{w}_2 = b + [g(\tilde{e}_2)/p]$, yields the period-one incentive compatible wage-effort schedule:

$$\tilde{w}_1 = \bar{v}_2 + \frac{g(\tilde{e}_1) - (1-p)g(\tilde{e}_2)}{p} \quad (4)$$

Although we have not yet explicitly determined neither the period-one and period-two profit maximising efficient wages, nor their associated effort levels, these variables will nevertheless have to comply with the supply conditions (2) and (4) respectively. Thus, given that $\bar{v}_2 = b$, any resulting wage profile will have to satisfy:

$$\Delta \tilde{w} \equiv \tilde{w}_2 - \tilde{w}_1 = \frac{(2-p)g(\tilde{e}_2) - g(\tilde{e}_1)}{p} \quad (5)$$

While we will infer more about the nature of the earnings profile under wage-only contracts later, we are presently unable to deduce whether the profile rises or falls with tenure. We can, however, still note the special case of constant effort where $g(\tilde{e}_1) = g(\tilde{e}_2) = g(\tilde{e})$ such that (5) reduces to $\Delta \tilde{w} = [(1-p)/p]g(\tilde{e}) > 0$. Intuitively, workers acquire rents on account of the firm's inability to perfectly monitor. The firm, however, can reduce these rents by offering lifetime contracts, irrespective of human capital considerations, that induce workers to queue up to gain access to the second period wage which exceeds their reservation utility. It should also be noted that whilst we preclude bonds being explicitly paid up front, younger workers may pay a type of entrance fee by accepting lower wages in their formative years.

4.2 Incentive Compatible Remuneration under PRP

Consider now a PRP contract in which total *actual* remuneration, \hat{w}_2^T , comprises both a fixed wage component, \bar{w}_2 , and a variable performance related output payment, $\theta_{it}f(\hat{e}_2)$, with λ reflecting the relative weight attributed to the performance element of the payment vis:

$$\hat{w}_2^T = (1-\lambda)\bar{w}_2 + \lambda\theta_{it}f(\hat{e}_2) \quad (6)$$

Workers are unable to observe the state of nature, θ_{it} , before they exert effort and so their decision regarding shirking will depend on total *expected* remuneration, $\hat{w}_2 = E\{\hat{w}_2^T\}$, which is given by:

$$\hat{w}_2 = (1 - \lambda)\bar{w}_2 + \lambda E\{\theta_{i2}f(\hat{e}_2)\} \quad (7)$$

where $E\{\theta_{i2}f(\hat{e}_2)\}$ denotes the worker's expected period-two revenue. A worker is labelled a shirker if he chooses to exert a level of effort \hat{e}_2 below the non-shirking level of effort \bar{e}_2 . In this case, the worker runs the risk of being fired with probability $p > 0$ such that the expected payoff to shirking is given by:

$$pb + (1 - p)[(1 - \lambda)\bar{w}_2 + \lambda E\{\theta_{i2}f(\bar{e}_2)\} - g(\bar{e}_2)] \quad (8)$$

Unlike his wage-only counterpart a PRP shirker will not typically exert zero effort but will instead apply the level of effort, $\bar{e}_2 \in [0, \hat{e}_2)$, that maximizes his utility payoff and which is defined implicitly from $\lambda \partial E\{\theta_{i2}f(\bar{e}_2)\} / \partial \bar{e}_2 = g'(\bar{e}_2)$. It follows from this that shirking is complete and effort is equal to zero in the special case when $\lambda = 0$, that is when a wage-only contract is offered as outlined in the previous section. Furthermore, shirking becomes less problematic as the firm introduces PRP (i.e. as λ increases) as shirking now punishes the worker through the lower performance related element that is part of the wider remuneration package. Indeed, as λ approaches unity and workers are paid entirely on the basis of output, the shirking problem all but disappears. There are, however, several reasons why wage-free contracts are rarely observed. If workers are risk averse, for example, then they will prefer contracts with at least a modicum of fixed wages in their remuneration package. And this preference will be further amplified if capital markets are imperfect thereby making consumption smoothing across time more difficult. We do not deal explicitly with the determination of λ in what follows but instead leave it as an externally determined parameter that is restricted on grounds of the above discussion.

The spot market/period-two *NSC* under PRP, denoted, NSC_2^{PRP} , takes the form:

$$\begin{aligned} & (1 - \lambda)\bar{w}_2 + \lambda E\{\theta_{i2}f(\hat{e}_2)\} - g(\hat{e}_2) \\ NSC_2^{PRP} \geq & pb + (1 - p)[(1 - \lambda)\bar{w}_2 + \lambda E\{\theta_{i2}f(\bar{e}_2)\} - g(\bar{e}_2)] \end{aligned} \quad (9)$$

Here, the fixed wage component of total period-two remuneration is given by \bar{w}_2 . Assuming the constraint holds as a strict equality then the above expression may be solved for the second period wage-effort schedule:

$$\bar{w}_2 = \frac{pb + g(\bar{e}_2) - \lambda[E\{\theta_{i2}f(\hat{e}_2)\} - (1 - p)[\lambda E\{\theta_{i2}f(\bar{e}_2)\} - g(\bar{e}_2)]]}{(1 - \lambda)p} \quad (10)$$

Note, however, that it is not the expected wage component per se that determines whether the worker supplies the required effort but rather the expected *total* remuneration, which comprises both a wage and performance related element. With the use of (10) we determine this latter as:

$$\hat{w}_2 = (1-\lambda)\bar{w}_2 + \lambda E\{\theta_{i2}f(\hat{e}_2)\} = b + \frac{1}{p}g(\hat{e}_2) + \left(1 - \frac{1}{p}\right)\lambda\Delta E_2 \quad (11)$$

where $\Delta E_2 = E\{\theta_{i2}f(\hat{e}_2)\} - [E\{\theta_{i2}f(\ddot{e}_2)\} - g(\ddot{e}_2)]$. It is apparent from (11) that an increase in $\lambda\Delta E_2$ reduces the period-two incentive compatible level of remuneration. Intuitively, a fall in the second term of ΔE_2 —i.e. $E\{\theta_{i2}f(\ddot{e}_2)\} - g(\ddot{e}_2)$ —reflects a fall in the value of period-two shirking thereby alleviating the required incentive compatible level of pay. pay value of shirking falls. A similar story may be told by an increase in the first term of ΔE_2 —i.e. $E\{\theta_{i2}f(\hat{e}_2)\}$. With $\lambda\Delta E_2$ representing the relative PRP gain of not shirking, it then follows that the incentive compatible level of remuneration falls as this rises.

To assure worker participation in the second period employment relationship, we assume that the utility derived from within the firm is at least equal to the utility of unemployment *vis.* $\hat{w}_2 - g(\hat{e}_2) \geq b$. From this and (11) we deduce that:

$$\hat{w}_2 - b - g(\hat{e}_2) = \left(1 - \frac{1}{p}\right)[\lambda\Delta E_2 - g(\hat{e}_2)] \geq 0 \quad (12)$$

Since period-two total expected remuneration, \hat{w}_2 , is anticipated in period-one it therefore enters the worker's lifetime period-one NSC, denoted NSC_1^{PRP} , which can be expressed as:¹³

$$NSC_1^{PRP} \geq \frac{(1-\lambda)\bar{w}_1 + \lambda E\{\theta_{i1}f(\hat{e}_1)\} - g(\hat{e}_1) + \hat{w}_2 - g(\hat{e}_2)}{p\beta + (1-p)[(1-\lambda)\bar{w}_1 + \lambda E\{\theta_{i1}f(\ddot{e}_1)\} - g(\ddot{e}_1) + \hat{w}_2 - g(\hat{e}_2)]} \quad (13)$$

Solving for the equilibrium period-one fixed wage element, \bar{w}_1 , yields the following wage-effort schedule:

$$\bar{w}_1 = \frac{p\beta + g(\hat{e}_1) - p[\hat{w}_2 - g(\hat{e}_2)] - \lambda[E\{\theta_{i1}f(\hat{e}_1)\} - (1-p)[E\{\theta_{i1}f(\ddot{e}_1)\} - g(\ddot{e}_1)]]}{(1-\lambda)p} \quad (14)$$

Given that $\bar{v}_2 = b$, the expected period-one total pay, \hat{w}_1 , for a non-shirker is thus expressed by:

¹³Note the right hand side of expression (13) represents the value of shirking in period-one. This value is maximized when the choice of shirking in period 1 satisfies $\lambda\partial E\{\theta_{i1}f(\ddot{e}_1)\}/\partial \ddot{e}_1 = g'(\ddot{e}_1)$. As the structure of shocks to productivity does not change overtime, it is straightforward by a comparison with the equivalent expression for period-two, set out in the paragraph following expression (8), that 'shirking' effort is intransient over periods such that $\ddot{e}_1 = \ddot{e}_2$.

$$\hat{w}_1 = (1 - \lambda)\bar{w}_1 + \lambda E\{\theta_{il}f(\hat{e}_1)\} = b + \left(\frac{1}{p}\right)g(\hat{e}_1) + \left(1 - \frac{1}{p}\right)[g(\hat{e}_2) - \lambda(\Delta E_2 - \Delta E_1)] \quad (15)$$

where $\Delta E_1 = E\{\theta_{il}f(\hat{e}_1)\} - [E\{\theta_{il}f(\tilde{e}_1)\} - g(\tilde{e}_1)]$ which denotes, when multiplied by λ , the instantaneous performance related gain in period-one from not shirking. In a direct reflection of the discussion following expression (11) of the analogous period-two gain—i.e. $\lambda\Delta E_2$ —it follows that the higher is this gain then the lower is the total level of remuneration required for incentive compatibility. An increase $\lambda\Delta E_2$, however, impacts positively on incentive compatible period-one remuneration as an increase in the gain from not shirking in period-two reduces incentive compatible remuneration in period-two, thereby reducing the ex ante period-one value of shirking to the worker. Thus, the firm responds by increasing period-one remuneration in order to achieve incentive compatibility.

Consider now the specific case where non-shirking effort is time invariant such that $g(\hat{e}_1) = g(\hat{e}_2) = g(\hat{e})$, which together with shirking effort $\tilde{e}_1 = \tilde{e}_2$ as shown by Footnote 14, implies $\Delta E_1 = \Delta E_2 = \Delta E$. This in turn implies from expression (15) that the first period remuneration under PRP is given by $\hat{w}_1 = \bar{v}_2 + g(\hat{e})$. Compare this to the first period wage in the wage-only contract as given by expression (4), when non-shirking effort is time invariant, that is $g(\tilde{e}) = g(\tilde{e}_1) = g(\tilde{e}_2)$, such that $\hat{w}_1 = \bar{v}_2 + g(\tilde{e})$. In words, if effort (and thus the cost of effort) is constant across the two time periods then there is no discernible difference between period-one remuneration across the two payment regimes, although period-two remuneration differences persist. Under either payment arrangement, period-one remuneration less effort cost is equal to outside option utility such that the worker is indifferent between employment and being dismissed. Nevertheless, the worker finds it beneficial to remain with the firm as wages and remuneration will rise in period-two. However, whilst the example is illustrative of the similarities across the two regimes, it is nevertheless overly restrictive as effort is not guaranteed to stay constant over the lifecycle as we demonstrate in Section 5.

Though we have not yet defined the efficiency remuneration in each period that maximizes the firm's profit, the payment profile must, as was the case for the wage-only arrangement, satisfy the supply conditions as given above. From expressions (11) and (15) we find that the total remuneration profile under PRP must satisfy:

$$\Delta \hat{w} \equiv \hat{w}_2 - \hat{w}_1 = \frac{1}{p}[(2 - p)g(\hat{e}_2) - g(\hat{e}_1) - (1 - p)\lambda(2\Delta E_2 - \Delta E_1)] \quad (16)$$

Note that an increase in the term $(1 - p)\lambda 2\Delta E_2$ affects the wage profile negatively in two (numerically identical) ways. First, an increase in this term reduces period-two remuneration as reflected in the discussion following expression (11); and second, it increases period-one remuneration as

reflected in the discussion following expression (15). Note also that the term $(1-p)\lambda\Delta E_1$ acts in the opposite manner.

4.3 Supply Side Comparisons Between Regimes

We observe by comparing (16) with (5) that it is theoretically possible for the remuneration gradient under PRP to be less steep than that under a wage-only contract. Note, however, that the converse is also possible and that which of the two regimes has the steepest profile will be dependent on the level of effort.

By way of illustration, consider the special, albeit unrealistic, case in which effort is invariant across both time and payment schedules implying a wage-only profile of $\Delta\tilde{w} = [(1-p)/p]g(\tilde{e}) > 0$, as shown previously, and a PRP profile of $\Delta\hat{w} = [(1-p)/p][g(\hat{e}) - \lambda\Delta E] \geq 0$. As $g(\hat{e}_1) = g(\hat{e}_2) = g(\tilde{e})$, it follows that $\Delta E_1 = \Delta E_2 \equiv \Delta E$.¹⁴ It is apparent that the wage-only profile is strictly upward sloping whilst the PRP profile is non-negative by implication of (9). By comparison of the expressions above the profile under PRP is flatter than under a wage-only contract when effort is invariant across payment schemes.

More generally, when effort levels are not time invariant, then we also note from a comparison of (5) and (16) that the shape of the earnings profile under pure salary and PRP schemes coalesce when either the extent of PRP or the impact of shirking on the PRP component of remuneration approach zero (i.e. as $\lambda \rightarrow 0$ or $\Delta E_1 \rightarrow \Delta E_2 \rightarrow \Delta E \rightarrow 0$).

To make more accurate assessments about actual payment profiles, when effort levels vary both across time and payment schemes we must extend the analysis beyond simple supply side considerations.

5 DEMAND SIDE ISSUES AND EFFICIENCY PAY

The previous section utilized the non-shirking condition to define the supply of effort and its relation to pay. With a whole locus of incentive compatible effort and remuneration combinations, we are however only somewhat closer to determining what the actual levels of compensation and effort should be. To tie down the equilibrium values of earnings and effort we now introduce the demand side with firms setting 'efficiency compensation' to maximize their profits subject to the workers behaving according to their previous determined supply pay-effort schedule.¹⁵ For clarity of exposition, we distinguish individual revenue from the (expected) aggregate revenue, $Y_t = F(e_t N_t)$, of employing N_t identical workers in period t .

¹⁴Note from the discussion following expression (8) that $\tilde{e}_2 \in [0, \hat{e}_2)$ such that $\Delta E_2 = E\{\theta_{12}f(\tilde{e}_2)\} - [E\{\theta_{12}f(\hat{e}_2)\} - g(\tilde{e}_2)]$ is unambiguously positive.

¹⁵We use the term efficiency compensation/pay/earnings/remuneration rather than efficiency wage when compensation includes more than a wage-only element.

We assume that the firm cannot commit to period-two remuneration in period-one but instead is left to choose the period specific control variables - wages and employment—that maximize the period specific profit, $\Pi_t = F(e_t N_t) - w_t N_t$, where $t = 1, 2$.¹⁶ Such a maximisation problem yields the Solow (1979) condition for efficiency wages vis. $(\partial e / \partial w)(w/e) = 1 \iff w = e / (\partial e / \partial w)$. The Solow condition characterizes the demand side and as such will be used with the supply information of the previous section.

5.1 Wage-only Contracts

In this sub-section we investigate the wage-only equilibrium effort in each period by combining both supply and demand behaviour. We can then deduce both the efficiency wage outcome and effort and how these are profiled over time.

Totally differentiating the period-two wage-effort supply schedule, $\tilde{w}_2 = b + [g(\tilde{e}_2)/p]$, implies $d\tilde{e}_2/d\tilde{w}_2 = p/g'(\tilde{e}_2)$. By combining this with $w = e / (\partial e / \partial w)$ from the Solow (demand side) condition, we derive the profit maximising efficiency wage, \tilde{w}_2^* :

$$\tilde{w}_2^* = \frac{g'(\tilde{e}_2^*)\tilde{e}_2^*}{p} \quad (17)$$

Substituting (17) into $\tilde{w}_2 = b + g(\tilde{e}_2)/p$ yields an expression for the optimal period-two level of effort:

$$g'(\tilde{e}_2^*)\tilde{e}_2^* - g(\tilde{e}_2^*) = pb \quad (18)$$

We repeat the above exercise for period-one by totally differentiating the supply expression (4), yielding $d\tilde{e}_1/d\tilde{w}_1 = p/g'(\tilde{e}_1)$. Using the Solow condition, we can then solve for the period-one efficiency wage:

$$\tilde{w}_1^* = \frac{g'(\tilde{e}_1^*)\tilde{e}_1^*}{p} \quad (19)$$

Substituting (19) into the left hand side of the supply expression (4) and rearranging then yields the condition for period-one equilibrium effort:

$$g'(\tilde{e}_1^*)\tilde{e}_1^* - g(\tilde{e}_1^*) = pb - (1-p)g(\tilde{e}_2^*) \quad (20)$$

A series of results now follow:

¹⁶Thus the firm-worker relationship is of a Crawford (1988) type and facilitates the use of the Solow condition outlined below in both period 1 and period 2.

Proposition 1: (i) Contractual effort in the wage-only setting is higher in period-two than in period-one.

(ii) The wage profile in the wage-only setting is upward sloping.

Proof: See Appendix. ■

Thus, we have shown that predictions from the agency literature of an upward sloping wage profile are retained. Taking the two parts of the proposition together, it is worth noting that the upward sloping wage-only contract raises the worker's fear of losing his job in period-two, which is the period in which the worker is 'cashing in' on the firm worker relationship. It therefore makes sense for both effort and wages to exhibit an upward sloping time profile. Note however, that the result depends heavily on the assumption of constant probability of detection, and were this assumption relaxed there would no longer be the same certainty regarding the progression of effort and wages as is outlined in the above proposition.¹⁷

5.2 The PRP Case

We now return to the PRP case in which the firm uses total remuneration, rather than just the wage, to elicit effort. Thus, we again consider the case where the firm maximizes profit, $E\{\pi_t\} = E\{F(e_t N_t) - w_t^* N_t\}$, in each of the two periods.

Using expression (11), the total remuneration-effort period-two (supply) schedule, yields:

$$\frac{\partial \hat{e}_2}{\partial \hat{w}_2} = \frac{p}{g'(\hat{e}_2) - \lambda(1-p)E\{\theta_{i2}f'(\hat{e}_2)\}} \quad (21)$$

Combining this with the Solow (demand) condition implies the optimal level of 'efficiency pay':

$$\hat{w}_2^* = \frac{\hat{e}_2^*}{p} [g'(\hat{e}_2^*) - \lambda(1-p)E\{\theta_{i2}f'(\hat{e}_2^*)\}] \quad (22)$$

By using (11) and (22), we are able to determine the profit maximising equilibrium level of effort:

¹⁷As is standard in the literature we thus preclude the possibility that effort affects the probability of detection. With effort of the non-shirkers affecting the probability of detection of the shirkers matters are more complex. In order to simplify the analysis we therefore assume that the firm or supervisor that monitors can detect shirkers at a fixed probability.

$$g'(\hat{e}_2^*)\hat{e}_2^* - g(\hat{e}_2^*) = pb + (1-p)\lambda [E\{\theta_{i2}f'(\hat{e}_2^*)\}\hat{e}_2^* - \Delta E_2] \quad (23)$$

We can now draw inferences about the effort profile in the PRP case:

Proposition 2: Period-two effort is lower under the PRP setting than the wage-only setting.

Proof: See Appendix. ■

As effort and remunerations are inextricably linked, inferences about second period differences across the two payment regimes can now also be drawn:

Proposition 3: Period-two remuneration is greater in the wage-only setting than in the PRP setting vis. $\tilde{w}_2^* > \hat{w}_2^*$.

Proof: Proposition 2 implies that $g(\tilde{e}_1^*) > g(\hat{e}_2^*)$. Given $\Delta E_2 > 0$ from footnote 15, Proposition 3 then follows by direct comparison of period-two wage remuneration under wage-only contracts and PRP from $\tilde{w}_2 = b + [g(\tilde{e}_2)/p]$ and expression (11) previously. ■

It appears from Propositions 1–3 that both second period remuneration and effort are higher in the wage-only case than the PRP case. This may reflect the particular agency issues that arise under the wage-only contract where wages are necessarily higher later on in the relationship. Thus, with remuneration higher later on in the wage-only arrangement it implies that period-two effort is also higher. Such a tendency is attenuated when the remuneration is a mix of wage-only payments and payment by performance.

To ascertain period-one effort under PRP, we return to expression (15), total differentiation of which implies:

$$\frac{\partial \hat{e}_1^*}{\partial \hat{w}_1^*} = \frac{p}{g'(\hat{e}_1^*) - \lambda(1-p)E\{\theta_{i1}f'(\hat{e}_1^*)\}} \quad (24)$$

Combining (24) with the Solow condition yields period-one efficiency remuneration under PRP:

$$\hat{w}_1^* = \frac{\hat{e}_1^*}{\hat{p}_1^*} [g'(\hat{e}_1^*) - \lambda(1-p)E\{\theta_{i1}f'(\hat{e}_1^*)\}] \quad (25)$$

Using (25) in conjunction with (15), we derive the period-one equilibrium level of effort:

$$g'(\hat{e}_1^*)\hat{e}_1^* - g(\hat{e}_1^*) = pb + (1-p)\lambda[(\Delta E_2 - \Delta E_1) + E\{\theta_{it}f'(\hat{e}_1^*)\}\hat{e}_1^*] - g(\hat{e}_2^*) \quad (26)$$

Proposition 4: Period-two PRP effort exceeds period-one PRP effort (i.e. $\hat{e}_2^* > \hat{e}_1^*$).

Proof: See Appendix. ■

Thus, both PRP and wage-only contracts yield upward sloping effort schedules. Note, however, that without imposing additional restrictions we are unable to draw conclusions about whether period-one effort under PRP exceeds period-one effort under a wage-only setting. Given that we do not know how period-one effort levels compare across the two remuneration arrangements, it also follows that period-one remunerations are not easily compared either. Nevertheless, when lifetime efforts under both regimes are equalized:

Proposition 5: Period-one wages are higher under PRP than under wage-only contract, when the net value of the wage-only contract does not exceed the PRP contract and when aggregate lifetime effort under the PRP contract is at least as high as the wage-only contract.

Proof: With the value of the wage-only contract being given as the sum of wages net of sum of efforts, $\tilde{w}_1^* - g(\tilde{e}_1^*) + \tilde{w}_2^* - g(\tilde{e}_2^*)$, across the two periods and the value of being employed under a PRP arrangement being the sum of total remuneration net of efforts, $\hat{w}_1^* - g(\hat{e}_1^*) + \hat{w}_2^* - g(\hat{e}_2^*)$, over the two periods, we have: $\hat{w}_1^* - g(\hat{e}_1^*) + \hat{w}_2^* - g(\hat{e}_2^*) \geq \tilde{w}_1^* - g(\tilde{e}_1^*) + \tilde{w}_2^* - g(\tilde{e}_2^*)$. With the aggregate lifetime effort condition $g(\hat{e}_1^*) + g(\hat{e}_2^*) \geq g(\tilde{e}_1^*) + g(\tilde{e}_2^*)$ it follows that: $\hat{w}_1^* + \hat{w}_2^* \geq \tilde{w}_1^* + \tilde{w}_2^*$. As Proposition 3 states that $\tilde{w}_2^* > \hat{w}_2^*$, it follows that $\hat{w}_1^* > \tilde{w}_1^*$. ■

The condition on which Proposition 5 relies, that the value of the two types of contract are the same, is of interest as it is the condition when the participation constraint binds in both instances. In order for the contracts to be both viable and to co-exist, their value will have to be equalized. With the additional restriction on aggregate effort from Proposition 5 in place, the wage profile is then flatter under PRP than under the wage-only setting.

That intuition for this result is straightforward. Period-two remuneration is anchored by the outcome in the spot market and it is relatively easier to satisfy the single period no-shirking condition under PRP as the firm has an additional instrument at its disposal. Like their wage-only counterparts, PRP shirkers run the risk of being detected and fired. Uniquely, however, even if they are lucky and manage to avoid detection, they suffer from a lower PRP element in their total remuneration package. As the firms trade off higher remuneration against higher effort and as it is relatively easier to satisfy the PRP

non-shirking constraint, it follows that end-period remuneration is lower under PRP than under wage-only setting. This in turn influences the period-two level of effort, which is lower under PRP than under the wage-only regime. Furthermore, there is now also an additional effect in the first period where, given the lower period-two remuneration under PRP, it is now relatively more difficult to satisfy the period-one *lifetime* non-shirking condition. The PRP firm responds to this difficulty by raising compensation, and thereby effort, in the first period.

It should be noted that our analysis is contingent on the restriction placed on lifetime effort levels. Relaxing this restriction could potentially reduce period-one PRP remuneration below period-one wage-only remuneration. If, for instance, PRP effort is sufficiently low then the PRP compensation schedule may be consistently below the 'wage-only' compensation schedule. It is then theoretically possible that the payment profile is steeper under PRP than under wage-only arrangements.¹⁸

6 FINAL COMMENTS

This paper has focused on the relationship between tenure-earnings profiles and the degree of worker 'equity' within an enterprise. We extend Lazear and Moore's (1984) thesis that the nature of the profile is primarily a reflection of agency considerations by focusing not only on those workers with zero or one hundred per cent equity (i.e. salaried and self-employed workers respectively), but also on those with a fractional level of equity *vis.* workers remunerated under some form of PRP.

The shape of the profile has important implications for labour market behaviour. If the slope is primarily a reflection of human capital considerations then it offers some clue as to the return to on-the-job training and educational investments. If agency considerations are paramount then it raises issues concerning the credibility of long-term employment contracts - firms may have an incentive to fire 'older' expensive, but no more productive, workers. A time-consistency problem may arise, with particular firms unable to recruit younger, less experienced applicants because of their inability to commit not to dismiss them in the future. If, however, the profile reflects

¹⁸Note that to derive Proposition 5 we imposed the further condition that the worker's value of the wage-only contract does not exceed the PRP value. Such a proviso would, of course, hold to equality in a market clearing, full-employment economy such that the assumption underlying Proposition 5 would be satisfied automatically. With unemployment, however, the values to the workers across payment regimes need not be equalized. Whether or not the value to the worker is greater under the wage-only or PRP contract is an open question that we do not tackle here. Instead, we simply observe that Proposition 5 would hold under the condition stated.

training then such an incentive-compatibility problem will not arise—older workers will be more productive *ceteris paribus*.

The nature of the earnings profile may also impinge upon quitting behaviour. Longer tenured ‘generally’ trained workers will have more manoeuvrability in the labour market than their otherwise similar ‘firm-specifically’ trained counterparts. But both types may have more options than those older workers whose market rents are primarily a reflection of agency considerations.

A simplifying assumption in the model concerns the state to which primary workers are (permanently) banished if found shirking vis. unemployment or lower paid secondary sector employment. Though scarring from dismissal due to underperformance may be substantial, in practice it may not be quite as severe as the complete scarring assumed here. Despite such concerns the assumption may be less restrictive than it first appears. Our focus is on the relative difference in the remuneration profiles faced by wage-only and PRP workers and assuming that the two types of workers have the same outside options, any scarring should affect both slopes identically. If, however, scarring differs across different types of contracts, then the relationship between worker equity and the tenure-earnings profile remains uncharted. We leave this possibility open for future discussions as it is beyond the scope of the current paper.

To the extent that our results support the view that the tenure-earnings profile reflects agency costs, they highlight important issues pertaining to the credibility of long-term employment contracts as employers may be tempted to replace tenured workers with less costly, but equally productive, novices. But the latter will not remain ‘young’ forever, and whether they will be inclined to work for a firm that is unable to guarantee them employment in their dotage is an open question.

Moreover, our findings may help to illuminate a hitherto neglected conduit for the transmission of productivity benefits under collective PRP schemes such as profit sharing: if capital markets are imperfect then the same tenure-earnings profile would inspire relatively less shirking under a profit-sharing as compared to a salaried contract on account of the lower degree of agency considerations that must be overcome. Alternatively, the same degree of effort may be obtained from risk averse workers *via* a flatter, and therefore less expensive, earnings profile.

APPENDIX

Proof of Proposition 1: Part (i):

Note that (18) and (20) jointly imply:

$$g'(\tilde{e}_2^*)\tilde{e}_2^* - g(\tilde{e}_2^*) = pb > g'(\tilde{e}_1^*)\tilde{e}_1^* - g(\tilde{e}_1^*) \quad (\text{A1})$$

Given that $g'(\tilde{e})\tilde{e} - g(\tilde{e})$ is increasing in effort, it follows that $\tilde{e}_1^* < \tilde{e}_2^*$.

Part (ii):

Note that (5) can be written as:

$$\Delta\tilde{w} \equiv \tilde{w}_2 - \tilde{w}_1 = \frac{(1-p)g(\tilde{e}_2) + [g(\tilde{e}_2) - g(\tilde{e}_1)]}{p} \quad (\text{A2})$$

From (A6) we have $g(\tilde{e}_2^*) > g(\tilde{e}_1^*)$. It therefore follows from (A2) that $\Delta\tilde{w}^* > 0$.

QED. ■

Proof of Proposition 2: Note that the concavity of the production function implies:

$$\begin{aligned} & \frac{E\{\theta_{12}f'(\hat{e}_2^*)\} - E\{\theta_{12}f'(\hat{e}_1)\}}{\hat{e}_2^*} > E\{\theta_{12}f'(\hat{e}_2^*)\} \\ & \Rightarrow \\ & E\{\theta_{12}f'(\hat{e}_2^*)\} - E\{\theta_{12}f'(\hat{e}_1)\} > \hat{e}_2^* E\{\theta_{12}f'(\hat{e}_2^*)\} \end{aligned}$$

Thus, with $g(\hat{e}_1) > 0$ we have:

$$\begin{aligned} & E\{\theta_{12}f'(\hat{e}_2^*)\} - E\{\theta_{12}f'(\hat{e}_1)\} + g(\hat{e}_1) > \hat{e}_2^* E\{\theta_{12}f'(\hat{e}_2^*)\} \\ & \Rightarrow \\ & \Delta E_2 > \hat{e}_2^* E\{\theta_{12}f'(\hat{e}_2^*)\} \\ & \Rightarrow \\ & \hat{e}_2^* E\{\theta_{12}f'(\hat{e}_2^*)\} - \Delta E_2 < 0 \end{aligned}$$

It, therefore, follows from expression (23) that under a PRP contract:

$$g'(\hat{e}_2^*)\hat{e}_2^* - g(\hat{e}_2^*) = pb + (1-p)\lambda[E\{\theta_{12}f'(\hat{e}_2^*)\}\hat{e}_2^* - \Delta E_2] < pb \quad (\text{A5})$$

Whereas (18) implies that under a wage-only contract:

$$g'(\tilde{e}_2^*)\tilde{e}_2^* - g(\tilde{e}_2^*) = pb \quad (\text{A6})$$

As the left hand sides of (A5) and (A6) are both increasing in effort, it follows that period-two effort is lower under a PRP contract than it is under a wage-only contract.

QED. ■

Proof of Proposition 4: The proof is given by contradiction: Assume contrary to proposition that $\hat{e}_2^* \leq \hat{e}_1^*$. Second and first period effort levels are given by expressions (23) and (26) respectively. Subtracting (23) from (26) implies:

$$\begin{aligned} & [g'(\hat{e}_1^*)\hat{e}_1^* - g(\hat{e}_1^*)] - [g'(\hat{e}_2^*)\hat{e}_2^* - g(\hat{e}_2^*)] \\ & = (1-p)\langle \lambda[(2\Delta E_2 - \Delta E_1) + E\{\theta_{i1}f'(\hat{e}_1^*)\}\hat{e}_1^* - E\{\theta_{i2}f'(\hat{e}_2^*)\}\hat{e}_2^*] - g(\hat{e}_2^*) \rangle \end{aligned} \quad (A7)$$

Given that the difference between the two terms on the left hand side of (A7) is positive when $\hat{e}_2^* \leq \hat{e}_1^*$, it follows that the right hand side must also be positive. This is rewritten below as:

$$(1-p)[\lambda\Delta E_2 - g(\hat{e}_2^*)] + (1-p)\lambda\langle [\Delta E_2 - E\{\theta_{i2}f'(\hat{e}_2^*)\}\hat{e}_2^*] - [\Delta E_1 - E\{\theta_{i1}f'(\hat{e}_1^*)\}\hat{e}_1^*] \rangle > 0 \quad (A8)$$

From (12) we have:

$$p[\hat{w}_2 - b - g(\hat{e}_2)] = (1-p)[g(\hat{e}_2) - \lambda\Delta E_2] \geq 0 \quad (A9)$$

Thus, we can re-express (A8) as:

$$-p[\hat{w}_2 - b - g(\hat{e}_2)] + (1-p)\lambda\langle [\Delta E_2 - E\{\theta_{i2}f'(\hat{e}_2^*)\}\hat{e}_2^*] - [\Delta E_1 - E\{\theta_{i1}f'(\hat{e}_1^*)\}\hat{e}_1^*] \rangle > 0 \quad (A10)$$

The first term of (A10) is non-positive by virtue of inequality in expression (A8). Concavity of production implies that if $\hat{e}_2^* \leq \hat{e}_1^*$ then:

$$\Delta E_1 - E\{\theta_{i1}f'(\hat{e}_1^*)\}\hat{e}_1^* > \Delta E_2 - E\{\theta_{i2}f'(\hat{e}_2^*)\}\hat{e}_2^* \quad (A11)$$

Thus the second term of (A10) is negative. A contradiction has thus been generated and so it must indeed be the case that $\hat{e}_2^* > \hat{e}_1^*$.

QED. ■

REFERENCES

- Adams, S. J. and J. S. Heywood. (2011). 'Does Deferred Compensation Increase Worker Effort?', *Manchester School*, Vol. 79, pp. 381–404.
- Akerlof, G. A. (1982). 'Labor Contracts as Partial Gift Exchange.', *Quarterly Journal of Economics*, Vol. 97, pp. 543–569.
- Akerlof, G. A. and Yellen, J. L. (1990). 'The Fair Wage–Effort Hypothesis and Unemployment', *Quarterly Journal of Economics*, Vol. 105, pp. 255–283.
- Altonji, J. G. and Williams, N. (2005). 'Do Wages Rise with Job Seniority? A Reassessment', *Industrial and Labor Relations Review*, Vol. 58, No. 3, pp. 370–397.
- Arulampalam, W., Gregg, P. and Gregory, M. (2001). 'Unemployment Scarring', *The Economic Journal*, Vol. 111, pp. 577–584.
- Basu, K. and Felkey, A. J. (2009). 'A Theory of Efficiency Wage with Multiple Unemployment Equilibria: How a Higher Minimum Wage Law Can Curb Unemployment', *Oxford Economic Papers*, Vol. 61, pp. 494–516.
- Becker, G. S. (1962). 'Investments in Human Capital: A Theoretical Analysis', *Journal of Political Economy*, Vol. 70, pp. 9–49.
- Ben-Porath, Y. (1967). 'The Production of Human Capital and the Life Cycle of Earnings', *Journal of Political Economy*, Vol. 75, pp. 352–365.
- Biewen, M. and Steffes, S. (2010). 'Unemployment Persistence: Is there Evidence for Stigma Effects?', *Economics Letters*, Vol. 106, pp. 188–190.

- Blinder, A. S. (ed.). (1990). *Paying for Productivity: A Look at the Evidence*, Washington, D.C., The Brookings Institution.
- Brown, S. and Sessions, J. G. (2006). 'Some Evidence on the Relationship Between Performance Related Pay and the Shape of the Experience-Earnings Profile', *Southern Economic Journal*, Vol. 72, No. 3, pp. 660–676.
- Bulow, J. I. and Summers, L. H. (1986). 'A Theory of Dual Labor Markets with Application to Industrial Policy, Discrimination, and Keynesian Unemployment', *Journal of Labor Economics*, Vol. 4, No. 3, pp. 376–414.
- Carmichael, L. (1990). 'Efficiency Wage Models of Unemployment: One View', *Economic Inquiry*, Vol. 28, No. 2, pp. 269–295.
- Crawford, V. P. (1988). 'Long-Term Relationships Governed by Short-Term Contracts', *American Economic Review*, Vol. 78, No. 3, pp. 485–499.
- Danthine, J. and Kurmann, A. (2006). 'Efficiency Wages Revisited: The Internal Reference Perspective', *Economics Letters*, Vol. 90, pp. 278–284.
- Demiralp, B. (2011). 'Occupational Self-Selection in a Labor Market with Moral hazard', *European Economic Review*, Vol. 55, pp. 497–519.
- Dohmen, T. (2004). 'Performance, Seniority, and Wages: Formal Salary Systems and Individual Earnings Profiles', *Labour Economics*, Vol. 11, No. 6, pp. 741–763.
- Dostie, B. (2005). 'Job Turnover and the Returns to Seniority', *Journal of Business and Economic Statistics*, Vol. 23, No. 2, pp. 192–199.
- Eliason, M. and Storrie, D. (2006). 'Lasting or Latent Scars? Swedish Evidence on the Long-Term Effects of Job Displacement', *Journal of Labor Economics*, Vol. 24, No. 4, pp. 831–856.
- Flabbi, L. and Ichino, A. (2001). 'Productivity, Seniority and Wages: New Evidence from Personnel Data', *Labour Economics*, Vol. 8, No. 3, pp. 359–387.
- Fuchs, V. (1981). 'Self-Employment and Labor Market Participation of Older Males', *Journal of Human Resources*, Vol. 18, pp. 339–357.
- Gibbons, R. (2005). 'Incentives Between Firms (and Within)', *Management Science*, Vol. 51, pp. 2–17.
- Gibbons R. and Waldmann M. (1999). 'A Theory of Wage and Promotion Dynamics inside Firm', *The Quarterly Journal of Economics*, Vol. 114, pp. 1321–1358.
- Gregg, P. (2001). 'The Impact of Youth Unemployment on Adult Unemployment in the NCDS', *Economic Journal*, Vol. 111, No. 475, pp. 626–653.
- Hilmer, M. J. (2011). 'Negative Returns to Seniority and Job Mobility Across the Program Quality Distribution: Are Top Public PhD-Granting Programs Different?', *American Economic Review*, Vol. 101, No. 3, pp. 466–470.
- Huck, S., Seltzer, A. J. and Wallace, B. (2011). 'Deferred Compensation in Multi-Period Labor Contracts: An Experimental Test of Lazear's Model', *American Economic Review*, Vol. 101, No. 2, pp. 819–843.
- Hutchens, R. M. (1987). 'A Test of Lazear's Theory of Delayed Payment Contracts', *Journal of Labor Economics*, Vol. 5, pp. 153–170.
- Hutchens, R. M. (1989). 'Seniority, Wages and Productivity: A Turbulent Decade', *Journal of Economic Perspectives*, Vol. 3, No. 4, pp. 49–64.
- Huttunen, K., Møen, J. and Salvanes, K. G. (2011). 'How destructive is Creative Destruction? Effects of Job Loss on Mobility, Withdrawal and Income', *Journal of the European Economic Association*, Vol. 9, No. 5, pp. 840–870.
- Lazear, E. P. (1979). 'Why is There Mandatory Retirement?', *Journal of Political Economy*, Vol. 87, pp. 1261–1284.
- Lazear, E. P. (1981). 'Agency, Earnings Profiles, Productivity and Hours Restrictions', *American Economic Review*, Vol. 71, pp. 606–620.

- Lazear, E. P. (1983). 'Pensions as Severance Pay,' in, Z. Bodie and J. Shoven (eds), *Financial Aspects of the U.S. Pension System*, Chicago, University of Chicago Press, pp. 57–85.
- Lazear, E. P. (2000a). 'The Future of Personnel Economics', *Economic Journal*, Vol. 110, F611–F639.
- Lazear, E. P. (2000b). 'The Power of Incentives', *American Economic Review, Papers and Proceedings*, Vol. 90, pp. 410–414.
- Lazear, E. P. and Moore, R. L. (1984). 'Incentives, Productivity and Labor Contracts', *Quarterly Journal of Economics*, Vol. 99, pp. 275–295.
- Lemieux, T., MacLeod, W. B. and Parent, D. (2009). 'Performance Pay and Wage Inequality', *Quarterly Journal of Economics*, Vol. 124, pp. 1–49.
- MacLeod, W. B. and Malcomson, J. (1998). 'Motivation and Markets', *American Economic Review*, Vol. 88, No. 3, pp. 388–411.
- Medoff, J. L. and Abraham, K. G. (1980). 'Experience, Performance and Earnings', *Quarterly Journal of Economics*, Vol. 95, pp. 703–736.
- Medoff, J. L. and Abraham, K. G. (1981). 'Are Those Paid More Really More Productive? The Case of Experience', *Journal of Human Resources*, Vol. 16, pp. 186–216.
- Mincer, J. (1958). 'Investment in Human Capital and Personal Income Distribution', *Journal of Political Economy*, Vol. 66, pp. 281–302.
- Ortín-Ángel, P. and Salas-Fumás, V. (1998). 'Agency-Theory and Internal-Labor-Market Explanations of Bonus Payments: Empirical Evidence from Spanish Firms', *Journal of Economics & Management Strategy*, Vol. 7 No. 4, pp. 573–613.
- Polachek, S. and Siebert, W. (1992). *The Economics of Earnings*, Cambridge, Cambridge University Press.
- Salop, S. (1979). 'A Model of the Natural Rate of Unemployment', *American Economic Review*, Vol. 74, pp. 433–444.
- Sessions, J. G. and Theodoropoulos, N. (2013). 'Tenure, Wage Profiles and Monitoring', *Research in Labor Economics*, Vol. 38, pp. 105–162.
- Shapiro, C. and Stiglitz, J. E. (1984). 'Equilibrium Unemployment as a Worker Discipline Device', *American Economic Review*, Vol. 74, pp. 433–444.
- Solow, R. (1979). 'Another Possible Source of Wage Stickiness', *Journal of Macroeconomics*, Vol. 1, pp. 79–82.
- Stiglitz, J. E. (1985). 'Equilibrium Wage Distributions', *Economic Journal*, Vol. 95, pp. 595–618.
- Strobl, E. and Walsh, F. (2007). 'Estimating the Shirking Model with Variable Effort', *Labour Economics*, Vol. 14, pp. 623–637.
- Walsh, F. (1999). 'A Multisector Model of Efficiency Wages', *Journal of Labor Economics*, Vol. 17, pp. 351–376.
- Weiss, Y. (1980). 'Job Queues and Layoffs in Labor Markets with Flexible Wages', *Journal of Political Economy*, Vol. 88, pp. 526–538.
- Weitzman, M. L. and Kruse, D. L. (1990). 'Profit Sharing and Productivity', in A. Blinder (ed), *Paying for Productivity: A Look at the Evidence*, Washington, D.C., The Brookings Institution.
- Wolpin, K. (1977). 'Education and Screening', *American Economic Review*, Vol. 67, pp. 949–958.
- Zwick, T. (2011). 'Seniority Wages and Establishment Characteristics', *Labour Economics*, Vol. 18, No. 6, pp. 853–861.